Sensors & Transducers, Vol. 159, Issue 11, November 2013, pp. 138-142

www.sensorsportal.com

Sensors & Transducers

© 2013 by IFSA

http://www.sensorsportal.com

RBF Neural Network Combined with Knowledge Mining Based on Environment Simulation Applied for Photovoltaic Generation Forecasting

[*] Dongxiao Niu, Ling Ji, Xiaomin Xu, Peng Wang

School of Economics and Management, North China Electric Power University,
Beijing, 102206, China

[*] Tel.: +86-10-5196-3828, fax: +86-10-8079-8531

[1] E-mail: niudx617@gmial.com

Received: 28 August 2013 /Accepted: 25 October 2013 /Published: 30 November 2013

Abstract: Photovoltaic generation forecasting is one of the main tasks of the planning and operation in power system. Especially with the development of mico-grid, relative study on renewable energy generation gain more and more concerns. In this paper, a short-term forecasting model combining knowledge and intelligent algorithm is developed for photovoltaic array generation. Self-organizing map (SOM) is proposed to extract the relative knowledge, and to choose the most similar history situation and efficient data for wind power forecasting with numerical weather prediction (NWP). The historical data is classified into several groups, though which we could find the similar days and excavate the hidden rules. According to the data reprocessing, the selected samples will improve the forecast accuracy radial basis function network (RBF) trained by the class of the forecasting day is adopted to forecast the photovoltaic output accordingly. A case study is conducted to verify the effectiveness and the accuracy. Compared with the conventional BP neural network, the forecasting results demonstrate the method proposed in this paper can gain better forecasting performance with higher accuracy. *Copyright © 2013 IFSA.*

Keywords: Knowledge mining, Self-organizing map, Radial basis function network, Photovoltaic generation, Forecast.

1. Introduction

With the climate change, energy resources depletion and cost increasing, the development and utilization of renewable energy (wind energy source, solar energy, geothermal energy and so on) has been the focus of the global attention at present. Since the theory of different renewable resource generation are various, and they all have the intrinsic characteristics of intermittent and uncertainty, the forecasting is an important and challenge job for the steady operation, effective schedule and system safety.

Recently, the home and abroad scholars focusing on load forecasting have adopted various theory and methods to study the photovoltaic generation forecasting. The forecasting method developed from the conventional regress and gray forecasting into the neural network, even the combined models with different methods, though which have improved the robustness and accuracy of forecasting gradually. In generally, the forecasting idea could be divided into two types, indirect method though solar radiation forecasting and direct method. In the former, main parameters like solar radiation, the area of array and the efficient of photovoltaic array can be converted into the output power of photovoltaic array [1]. Regression smoothing average model and neural network are the typical methods of the latter. Stanley *et al.* adopt the artificial neural network (ANN) to minimize the nonlinear correlation between the metrological parameters and energy generated by the PV system [2]. In paper [3], a new statistical short-term forecasting system for a gird-connected photovoltaic plant is presented, and it comprised thee modules composed of two numerical weather prediction models and an artificial neural network based model.

In the photovoltaic generation technology, the light is converted into electricity power though the Volta affected though the different semiconductor. Solar irradiation is the main influent factor. It represents the amount of solar irradiation on some area during certain time period, which is described with kWh/m^2. In fact, the climate conditions include the temperature, humidity, floating dust, cloud cover, and air pollution. Considering these climate conditions, Patrick *et al.* develop a high-resolution, direct-cloud-assimilating NWP based on the Weather and Research Forecasting model (WRF-CLDDA) [4]. The output power of photovoltaic generation system changes with the different solar irradiation on the surface of the earth.

Thus, the main goal of this paper is attempted to build an intelligent model combined with knowledge mining. By utilizing the properties of SOM and RBF, the algorithm for photovoltaic generation is proposed in this paper. The rest parts of this paper are organized as follows. Knowledge mining and SOM theory is briefly introduced in Section 2. The RBF neural network is described in Section 3. A case study is demonstrated and the performance of the proposed algorithm is discussed in Section 4. Conclusions are drawn in Section 5.

2. Knowledge Mining and SOM

Data mining is most commonly used in attempts to derive useful information and extract useful patterns from large data sets or database for solving specific issue [5]. The knowledge data base includes dominant and recessive, qualitative and quantitative, empirical and inferential knowledge, in addition, all kinds of knowledge could be analyzed automatically and be conducted though inductive reasoning. The gained potential patterns and rules could be served for forecasting decision, pattern recognition, fault diagnosis and production process optimization. Lots of data mining technologies have been developed and applied in control problem and classification. Self-organization mapping (SOM), firstly proposed by Kohonen in 1981, is a neural network with unsupervised learning [6]. It has certain topological structure which is adjusted through the input information, and the pattern recognition is completed by the synergy among multiple neurons [7-8]. The special idea of SOM is that there is no need to initialize the cluster center or guidance information, and that the weight information of neurons is self-adaptively adjusted by input data.

Generally, there are two layers in SOM neural network: the first layer for the sample input, and the number of neurons is equal to the dimension of samples; the second layer is competition layer, in which each neuron represents a class, therefore the number of the neuron in competition layer is the same with the clustering categories. Two-dimension is the most common form of the neurons' arrangement in competition [9-11]. The basic principle of SOM is described as following.

The weight with the closest match to the presented input pattern is the winner neuron or the best matching unit (BMU).

1) Competition process.

Let P be the input samples with m dimensions $P=\left[p_1,p_2,...,p_m\right]^T$, m is the number of the input neurons. The number of the neurons in competition layer with two dimension is N (N = n*n). The connection weight between the input layer and the competition layer are denoted as $W=\left[W_1^T,W_2^T,...,W_N^T\right]^T$, and $W_i^T=\left[w_{i1},w_{i2},...,w_{im}\right]$.

Calculate the inner product n of input vector and the connection weight:

$$n=\left[n_1,n_2,...,n_N\right]=WP=\left[W_1^TP,W_2^TP,...,W_N^TP\right]^T \tag{1}$$

Select the winner neuron in the competition process or the best matching unit (BMU). The basic rule is that the larger the inner product is, the closer the neuron to the input vector, which indicates the neuron matches to the presented input pattern. The winning formula is as follows:

$$a=compet\left(n\right)=\begin{cases}1, & i=i^*\\0, & i\neq i^*\end{cases}, \tag{2}$$

where $n_i^*\geq n_i$, $\forall i$, and $i^*\leq i$, $\forall n_i=n_i^*$. Finally, only one neuron wins, thus the result is similar as $a=\left[0,...1...,0\right]$, and i^* is the winner neuron.

2) Learning process.

SOM neural network is arranged according to the two-dimensional structure, each neuron has promoting effect to the neighboring neurons, on the contrast, inhibition effect to those far away, as shown in Fig. 1. Where d_{ij} denotes the distance between the neuron and its neighbor, Δw_{ij} is the change of the connection weight. It indicated that the neurons within the certain scope are promoted with weights increasing; while the neurons out of the scope are inhibited, and the weigh would be reduced.

In SOM neural network, the weight is adjusted according to Kohonen learning rules. The main idea is to make the winner neuron and its neighbors closer to the input sample by modifying the weights.

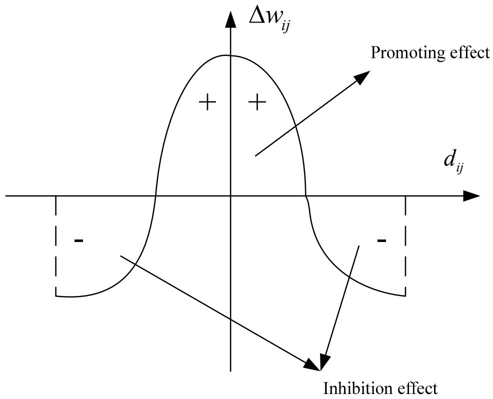


Fig. 1. The effect of a neuron to its neighbors.

The formula is expressed as:

$$w_i(t) = w_i(t-1) + \alpha(t)(P_i(t) - w_i(t-1)), \quad (3)$$

$$\alpha(t) = \alpha(0) * \left(1 - \frac{t}{T}\right), \quad (4)$$

$$N(t) = N(0) * \left(1 - \frac{t}{T}\right), \quad (5)$$

where $w_i(t)$ and $w_i(t-1)$ represent the weight vector of neuron i at time t and $t-1$ respectively. $P(t)$ is the input sample at time t. And $\alpha(t)$ means the learning rate at time t, ranging in [0, 1]. At the beginning, the learning rate is the largest, and the value of $\alpha(t)$ is decreasing with the training. Since the weight shock of the neurons may occur during the training process, the learning rate need to be reduced gradually. Training neighborhood $N(t)$ represents the neurons around the winner neuron whose weights will be adjusted. In the initial training, the scope of neighborhood is largest, which provides more neurons learning opportunity. With the increasing of training, each neuron represents its own categories and its neighborhood will be less.

3. RBF Neural Network

In this paper, the main variables affecting the generation power of photovoltaic array including the maximum temperature (T_{max}), the minimum temperature (T_{min}), solar radiation intensity (G), average humidity (H), cloud cover (C) and average wind (W) are consist of the environment characteristic space V. RBF neural network is built to approximate the complicated non-linear relationship between the input characteristic space V and photovoltaic the output power P.

RBF neural network take the advantages of simple topological structure, good smoothness, quick convergence speed and no local minima, and can approach any nonlinear system. The structure with three layers is illustrated in Fig. 2, where $\Phi(\cdot)$ represents the RBF. The first layer is consisted of several perception units. The second layer connected with network with outside environment is called hidden layer, executing the non-linear exchange for feature extraction. Compared with conventional back propagation (BP) neural network, the number of the hidden nodes in RBF neural network can be determined in light of requirement. In addition, it overcomes the disadvantages of BP with slow convergence speed and minimum local [12-14].

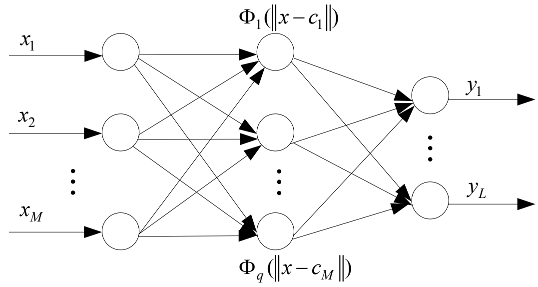


Fig. 2. The structure of RBF neural network.

Suppose the input of the network x with M dimension, the output y with L dimension, and the non-linear mapping $x \to u_i(x)$ is realized in the hidden layer.

Gaussian function is adopted as the activation function of the hidden nodes in RBF neural network. Thus, the output of the i^{th} node would be expressed as:

$$u_i = \exp\left[-\frac{(x-c_i)^T(x-c_i)}{2\sigma_i^2}\right] \; i = 1,2,...,L \quad (6)$$

The linear mapping from hidden layer to output layer could be described as:

$$y_k = \sum_{i=1}^{q}(w_{ki}u_i) - \theta_k \; i = 1,2,...,L, \quad (7)$$

where $x = (x_1, x_2, ...x_M)^T$ denotes the input samples; q is the number of the nodes in hidden layer; u_i is the output of the i^{th} node in hidden layer; σ_i is the standardized constant of the i^{th} node in the hidden layer; y_k is the output of the k^{th} node in output layer; w_{ki} is the weight coefficient from hidden layer to output layer; θ_k denotes the threshold of nodes in output layer; c_i is the center vector of the Gaussian function of hidden layer nodes, and this vector is a column vector with the same input sample x, $c = (c_{i1}, c_{i2}, ..., c_{iM})^T$.

The essential problem of RBF neural network design is to determine the number of the nodes in hidden layer and the center and the width of the relative radial basis function though learning with the purpose of minimum the forecast error and granting the generalization ability.

4. Case Study

In this paper, the photovoltaic array in the Northwest of China is taken as the case study to verify the proposed method. The time horizon of output power is 15 min, and the relative climate data are taken as the input variables. Suppose part of the input variables as the training set, and the rest as the test set. The feature vector V is consist of the maximum temperature (T_{max}), the minimum temperature (T_{min}), solar radiation intensity (G), average humidity (H), cloud cover (C) and average wind speed (W). Select two days *a* (sunny) and *b* (cloudy), and find their similar days though SOM neural network. The subsamples are utilized to train the RBF neural network respectively. In addition, commonly used standard BP neural network is also employed to forecast the photovoltaic generation with the same data. The samples in the same class with forecasting day are selected to train the ESN network for the forecasting day. In order to test the performance of ESN, BP neural network which has been widely used in load forecasting is also applied for the same task. Commonly used error evaluation indexes, including mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE), are used to discuss the performance of different forecasting methods:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{8}$$

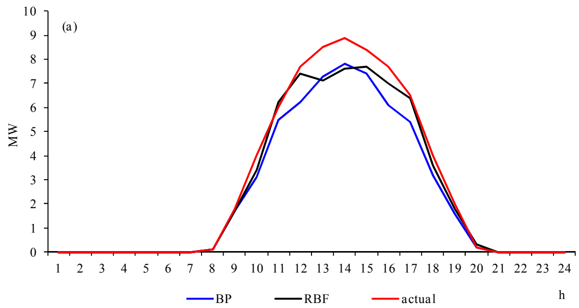
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}, \tag{9}$$

where $\hat{y}_i$ is the forecasting value and y_i is the actual value, N is the number of samples. The performances of different forecasting methods are shown in Table 1.

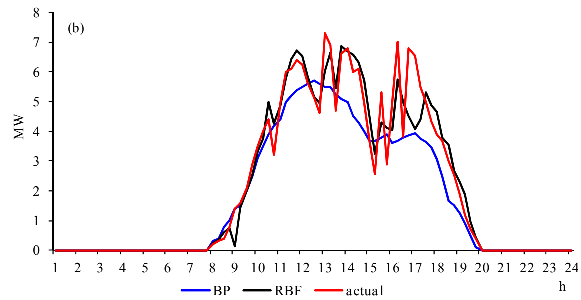
Table. 1. The performance of different forecasting methods.

Model Classification	RMSE (MW)		MAPE (%)	
	RBF	BP	RBF	BP
Sunny model	0.69	1.32	6.23	8.54
Cloudy model	0.85	2.49	13.27	21.65

Fig. 3 shows the forecasting results of different methods and the actual wind power output. It indicated that the overall trend of forecasting power is in accordance with the actual situation. However, the performance of RBF is obviously greater than BP model. The MAPE and RMSE of RBF under sunny day are 6.23 % and 0.69 MW respectively, lower than those of BP 8.54 % and 1.32 MW. While as to the cloudy day, the RMSE and MAPE of RBF forecasting is 0.85 MW and 13.27 %, and those of BP is 2.49 and 21.65 %, respectively. This indicated that both of the methods gain better accuracy when solar radiation intensity is stable.



(a)



(b)

Fig. 3. The forecasting and actual results of photovoltaic array under sunny day (a), and cloudy day (b).

5. Conclusions

Photovoltaic forecasting is an important tool for managing the inherent variability and uncertainty in photovoltaic generation. Increasing the accuracy of forecasting can help to reduce the likelihood of an unexpected gap between scheduled and actual photovoltaic power generation, which can be extremely helpful for operators of power systems. In this study, we developed a database by using historical meteorological environment and power output data. SOM neural network as a knowledge mining technology is employed to discover and extract the rules. The classified samples are taken as the input of echo state network to train the structure of the network respectively. Through clustering the historical data in to several classes, the approach would provide valid data for ESN forecasting. The results of case study demonstrated the successful use of the proposed method, which performs better than BP. The accuracy of prediction has been improved. However, the database is static in this study, which means the information of new samples will not be added into the knowledge pool automatically. And if the database is small, it may not cover comprehensive situation. Thus, it will be proper for the photovoltaic power station with long operation period.

Acknowledgements

This work is supported in part by NSFC under Grant No. 71071052 and "the Fundamental Research Funds for the Central Universities" under Grant No. 12QX23. In addition, it is also supported by China Scholarship Council (CSC).

References

[1]. J. Shi, W. J. Lee, Y. Q. Liu, Y. P. Yang, P. Wang, Forecasting power output of photovoltaic systems based on weather classification, *IEEE Transactions on Industry Applications*, 48, 3, 2012, pp. 1064-1069.

[2]. S. K. H. Chow, E. W. M. Lee, D. H. W. Li, Short-term prediction of photovoltaic energy generation by intelligent approach, *Energy and Buildings*, 55, 2012, pp. 660-667.

[3]. E. G. G. Villena, M. L. S. Pedro, Z. A. Enrique, J. Z. S. Pedro, Short-term power forecasting system for photovoltaic plants, *Renewable Energy*, 44, 2012, pp. 311-317.

[4]. P. Mathiesen, C. Collier, J. Klesissl, A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting, *Solar Energy*, 92, 2013, pp. 47-61.

[5]. C. H. Chen, A. F. Li, Y. C. Lee, A fuzzy coherent rule mining algorithm, *Applied Soft Computing*, 13, 2013, pp. 3422-3428.

[6]. Z. H. Guo, Y. Dong, J. Z. Wang and H. Y. Lu, The forecasting procedure for long-term wind speed in the Zhangye area, *Mathematical Problems in Engineering*, Vol. 2010, Article ID 684742.

[7]. Z. S. Wang, S. R. Bian, Y. Liu, The load characteristics classification and synthesis of substations in large area power grid, *International Journal of Electrical Power & Energy System*, 48, 2013, pp. 71-82.

[8]. L. Hemandez, C. Baladron, J. M. Aguiar, Classification and cluster of electricity demand patterns in industrial parks, *Energies*, 5, 12, 2012, pp. 5215-5228.

[9]. A. Seret, T. Verbraken, S. Versailles, B. Baesens, A new SOM-based method for profile generation: Theory and an application in direct marketing, *European Journal of Operational Research*, 220, 2012, pp. 199-209.

[10]. W. C. Liao, C. C. Hsu, A self-organizing map for transactional data and the related categorical domain, *Applied Soft Computing*, 12, 2012, pp. 3141–3157.

[11]. P. Juntunen, M. Liukkonen, M. Lehtola, Y. Hiltunen, Cluster analysis by self-organizing maps: An application to the modelling of water quality in a treatment process, *Applied Soft Computing*, 13, 2013, pp. 3191-3196.

[12]. W. Y. Zhang, J. J. Wang, J. Z. Wang, Z. B. Zhao, M. Tian, Short-term wind speed forecasting based on a hybrid model, *Applied Soft Computing*, 13, 2013, pp. 3225-3233.

[13]. J. D. Wu, J. C. Liu, A forecasting system for car fuel consumption using a radial basis function neural network, *Expert Systems with Applications*, 39, 2012, pp. 1883-1888.

[14]. M. Gan, H. Peng, Stability analysis of RBF network-based state-dependent autoregressive model for nonlinear time series, *Applied Soft Computing*, 12, 2012, pp. 174-181.

2013 Copyright ©, International Frequency Sensor Association (IFSA). All rights reserved.
(http://www.sensorsportal.com)